Evaluation of a telepathology system between Boston (USA) and Dijon (France): Glass slides versus telediagnostic TV-Monitor.

F.A. Allaert[1] , D. Weinberg[2], P. Dusserre[3] P. J. Yvon[4] L. Dusserre[1], P. Cotran.[2] Dpt of medical informatics university of Dijon. France. 2) Dpt of pathology Brigham and Women's Hospital Harvard Medical School. USA.3) Centre de pathologie Dijon. 4) Resintel France.

The objective of this work was to compare diagnoses achieved through the traditional methods of current pathology practice versus diagnoses achieved through a selection of image on a telediagnostic TV-monitor. The Kappa coefficient between the two protocols of k=0.26 SE(k)=0.06 z=k/SE(k)=4,3 (p<0.001) allows us to conclude that there is a good reliability between video and glass slide diagnoses.

INTRODUCTION

Telemedecine systems and specially telepathology systems require evaluation (1, 2, 3). In France a telepathology network developed by Resintel is now in daily use, bringing to isolated medical doctors help in making their diagnosis . This has proved of great usefulness in interactive situations, when two physicians converse by telephone, showing each other the part of the lesion they find necessary to discuss (4). In the future this interactive process could be replaced by an electronic mailing service in which a limited number of images are taken from the slide, but this protocol needs to be assessed in comparison with traditional microscopic diagnosis when the pathologist has the possibility of screening the entire slide.

The objective of this work was to compare diagnoses achieved through the traditional methods of current pathology practice versus diagnoses achieved through a selection of image on a telediagnostic TV-monitor. The department of pathology of the Brigham and Women's Hospital was choosen as the reference site.

MATERIALS AND METHODS

Case material

200 sequential random cases of routine surgical pathology were selected from the files of cases seen at Centre de Pathologie (Dijon, France), and a single representative slide was selected from each case. The distribution of cases according to organ system are listed in the following table.

Organ/system	Number
Gastro-intestinal	42
Gynecological	35
Skin	35
Head and neck	19
Breast	13
Soft tissue	9
Lung	6
Bladder	6
Lymph node	6
Synovium, tendon	6
Penic, testic, cord	6
Thyroid	6
Prostate	5
Kidney	3
Blood vessel	3
Bone	1
Total	200

Among the cases, there were 160 benign and 40 malignant lesions. One pathologist was responsible for slide selection, as well as selection of microscopic fields for image capture (see below). The clinical history consisted of that which was provided by the clinician at the time of submission of the tissue specimen. The cases were divided into four groups (A,B,C,D) of 50 cases, and numbered

0195-4210/95/$5.00

sequentially (1-200) for the purposes of labelling the video image files. An additionnal 50 cases were selected for the purposes of training. The glass slides were later assigned random case number within each group of 50 slides, and transported to Brigham and Women's Hospital (BWH), along with paper copies of the clinical histories. Investigators at BWH were not aware of the original diagnoses or the original case numbers of the glass slides.

Image capture

A single representative glass slide from each of 200 sequential random surgical pathology cases submitted to Centre de Pathologie was selected for the study. Images of representative microscopic fields were captured using the Transpath workstation at the discretion of the pathologist, 5-12 images per case, sampling at low (100X), intermediate (200X), and high (400X) magnifications, as seemed appropriate to render a diagnosis. The Transpath workstation consists of a standard light microscope (Leica, DMRB) equipped with a three-chip RGB color video camera (sony DXC M7), capable of capturing 720 X 600 pixels, linked to a digital image capture board (Avis CCIR 601, 702 X 576 pixels, 24 bit color, 4.2.2 standard) contained in a desktop personal computer powered by a 386 microprocessor. Image storage and retrieval is managed using the system software (VT Com TRIBUN company, Paris, France). 24 bit color image files were compressed using standard JPEG hardware compression, stored on the 200 MB hard drive, and later transfered along the clinical information text files to CD-ROM (Image Directe, Paris, France).

Image Display

The CD-ROM images were displayed at BWH using a MacIntosh Quadra 800 computer , equipped with a 24 bit color display card , a 14 inch monitor, a 17 inch color monitor and an internal double speed CD-ROM drive. Computer software included System 7.1 Quicktime (v. 1.6), and image viewing software authored by TRIBUN (Paris). The images were displayed using the 4.1.1 standard, as 768 X 574 pixels. The viewing software allows the selection of cases from a file listing the cases sequentially, with all of the images pertaining to the chosen case displayed as a gallery on the 14 inch monitor. Images selected from the gallery are displayed a full size on the 17 inch monitor. A window containing the clinical history is presented on the screen along with the image gallery. The pathologists at BWH involved in the study were trained to use the viewing workstation using the training set of 50 images, and did not begin the study until they were familiar with the system and comfortable viewing the still images.

Study Protocol

Each pathologist (A,B,C,D) was asked to view the cases stored on CD-ROM in five sessions of ten cases per session, with no more than one session per day, over a period of three weeks. Cases were selected sequentially by the viewing pathologist from the list presented on the computer monitor, and a gallery of all of the images, in reduced size, was presented as a "gallery" on the 14 inch monitor, along with the clinical history. Each image in the gallery was selected using the mouse for display at full size in the 17 inch monitor. Upon completion of viewing of the case, data sheet was completed which included the case number, diagnosis, certainty of diagnosis, and reason for each session of ten cases

was also recorded. All data sheets were submitted to one of the investigators (DSW) for recording of the data and storage until completion of the study.

Three to four weeks following completion of the viewing of the CD-ROM images, the glass slides from the corresponding cases, at random numbered, were distributed to each of the four pathologists. The pathologists were not told that they would be seeing the same group of fifty cases each had viewed on the computer, but only that the cases had been randomized. They were instructed to review the glass slides in groups of ten cases per viewing session. Again, the diagnostic information, including the degree of uncertainty and reasons for uncertainty, as well as the time required to view each group of ten cases, were recorded and submitted. For this study, the pathologist was allowed to consult books and other information to arrive at a diagnosis, but could not obtain consultation from colleagues.

For each case, the pathologist was required to record the degree of diagnostic certainty, ranging from "absolutely correct" (level 1) to "uncertain (educated guess)" (level 5). For all levels of certainty other than 1, the possible reasons for uncertainty were also recorded.

Data Analysis

At the conclusion of the study, the code was broken, and the diagnoses performed at the original institution in Dijon, and at BWH by viewing the CD-ROM images (CD) and the corresponding glass slides (GS) were compared. Diagnoses were recorded as : (a) completely concordant ; (b) discordant for language only ; (c) discordant, clinically not important ; and (d) discordant, clinically important. A discordance was considered clinically important if there would be any difference in clinical treatment or follow-up required based on this error in diagnosis. Overall, cases belonging to groups "a" and "b" were grouped as concordant.

All cases having a level "c" or "d" discordance were reviewed in order to determine the correct diagnosis. The glass slide from each such case were shown to the daily meeting of the surgical pathology staff at BWH, consisting of approximately 10 pathologists who were not involved in the study, and a consensus diagnosis was achieved. All cases requiring specialty pathology consultation were reviewed by appropriate specialty groups at BWH (e.g., dermatopathology). In this manner, the "correct" diagnosis was assigned a to each case in which there was any discordance among the three diagnoses (Res, CD, GS). The clinical significance of the discordance was similarly determined, by consensus. In addition, the CD-ROM images from all cases in which the CD an GS diagnoses were discordant were reviewed by one of the study members to determine the source of the error (i.e., interpretation, sampling, field selection, or image quality).

The Kappa statistic of Cohen was used as an indicator of observer agreement (5). This statistic is a chance-corrected measure of agreement and ranges up to 1, depending of the strength of agreement. If the statistic takes the value of zero, this indicates no agreement, negative values indicate disagreement, and positive values indicate agreement. To test the hypothesis of no observer agreement while adjusting for chance agreement, the statistic Z=K/STD(K) is compared to the standard normal distribution of Z for a two-tailed test.

RESULTS

The following table shows the intraobserver correlation of correct and incorrect diagnoses.

Glass Slide vs TV Monitor	Correct diag.	Incorrect diag.
Correct diagnosis	173	18
Incorrect diagnosis	4	5

The agreement rate is 86% if we just take into account the correct diagnoses or 87,5% if we consider also 2 cases among the incorrect one for which the diagnosis performed by video image and glass slide is wrong but exactly the same.

The Kappa coefficient between the two methods of diagnosis is $k=0.26$ $SE(k)=0.06$ $z=k/SE(k)=4,3$ ($p<0.001$) indicating that there is a good agreement between video and glass slide diagnoses.

DISCUSSION

There are relatively few studies which have addressed the accuracy of telepathology, compared to the larger number of such studies in teleradiology (7). Remote frozen section diagnosis during surgery may represent an important application of telepathology, and several studies have examined the accuracy of telepathology for this purpose. Nordrum et al. (8), using a combined dynamic and static telepathology system in Norway, were able to make a correct diagnosis in all 17 cases attempted. In Switzerland, Oberholzer et al (9), using transmitted still images to attempt frozen section in 16 cases, found a sensitivity of 50% and a specificity of 100% for a diagnosis of malignancy. Becker et al.(10), using the Telmed System to transmit static images of neurosurgical frozen sections to the Armed Forces Institute of Pathology (Washington D.C.) over standard telephones lines, realized 87% of concordance with the permanent section diagnosis. More recently, Weinstein and col (13) have examined the accuracy of a static image pathology network with regard to diagnosis of 126 routine surgical cases, and found discordance in 6% of cases. Clearly, studies involving larger number of cases are needed to further test the accuracy of diagnostic telepathology.

Differences in diagnostic practice make the determination of diagnostic accuracy more difficult for pathology than for radiology. In most teleradiology studies, diagnosis is limited to the determination of the presence or absence of a lesion, and thus one can determine the sensitivity and specificity of the binary decision by the radiologist (7,12). Weinstein and col (13) have successfully employed this approach in comparing the accuracy of frozen section diagnosis of breast lesions (benign vs. malignant) using standard microscopy and video microscopy. However, there is far greater range of diagnostic possibilities in general surgical pathology, and allowance must be made for individual differences in descriptive and diagnostic language.

Our study has several limitations which should be considered in interpreting the results. First, the types of cases which were used more frequently for subspecialty consultation, and the technical demands may be different than for "routine" pathology. Second, the pathologist viewing the CD-ROM images were prevented from seeking consultation from colleagues, or from reffering cases to subspecialists. This approach differs greatly from current practice, in which subspecialty consultation is frequently sought. Third, the pathologist was prevented from discussing the case with the "referring" pathologist and could not request additional images or clinical history. Therefore, it is likely that the accuracy of telepathology as actually practiced would be greater

than our results indicate. Moreover, examination of individual performance indicated that one of the four pathologists was responsible for almost half the total discordances and errors based on CD diagnosis. This pathologist performed as well as the others in performing glass slides diagnosis, indicating that there may be great individual differences in ability to adapt to the viewing of video images. Therefore, some pathologists may requires greater amounts of training than others to acquire the necessary visual skills.

CONCLUSION

This study demonstates that pathologic diagnosis based on passively-acquired still images correlates well with glass slide diagnosis and that telepathology is useful to bring expertise to pathologists who are working in such conditions they can't seek consultation from colleagues or refer cases to subspecialists. However, it is important to point out that great care most be taken in the selection of microscopic field and that interaction with the referring site and access to consultation with colleagues might improve the accuracy of diagnosis. It will be also necessary to provide adequate training of pathologists for telepathology, as some of the skills required may be different from those used for standard microscopy. More research must be performed to study the optimal technical requirements for telepathology and to develop standards of practice. Until technical and quality assurance standards exist, it will be necessary for each pathology site to establish the accuracy of its diagnostic services.

Références

1) S. Olsson. Evaluation of the swedish telepathology trial. Second european meeting on telepathology.Paris 9-10 june 1994.

2) Allaert F.A , P. Dusserre, L. Dusserre. Telediagnosis through a digital picture network: technical aspect and quality assesment. Ninth Annual International Symposium on Computerization of Medical records combined with the North American Conference on Patients Cards. San Antonio 22-25 april 1993, Texas, USA.Proceedings p 136 -140.

3) A.M. Elsayed. Technical requirements for a telepathology system. Second european meeting on telepathology.Paris 9-10 june 1994.

4) Dusserre P., Allaert F.A., Dusserre L. Martin E. Télédiagnosis through a digital picture network: The french pathologists' experience. Proceedings of MIE 93 . XI international congress of the European Federation for medical informatics, Jerusalem april 18-22.Freund Publishing House Ltd, p 428 - 428.

5) Cohen J. Weighted kappa: normal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70: 213-220, 1968.

6). A. Taube. To judge the judges: Kappa, ROC or What? Second european meeting on telepathology.Paris 9-10 june 1994.

7) Goldberg M.A., Rosenthal D.I., Chew F.S., Blickman J.G., Miller S.W., Mueller P.R.. New High-resolution teleradiology system: prospective study of diagnostic accuracy in 685 transmitted clinical cases. Radiology 186: 429-434, 1993.

8) Nordrum I., Engum M;, Rinde E.: Remote frozen section service: a telepathology project in northern norway. hum Pathol 22: 514-518,1991.

9) Oberholzer M., Fisher H-R, Christen H.: Telepathology with an integrated services digital network. A nex tool for image transfer in surgical pathology. A preliminary report. Hum Pathol. 24: 1078-1085, 1993.

10) Becker R.L., Specht C.S., Jones R. Use of remote videomicroscopy (telepathology) as an adjunct to neurosurgical frozen section consultation. Hum Pathol 24: 909-911,1993.

11)Bhattacharyya A., Halliday B.E., Weinsten R.S.: Telepathology: clinical assessment of an international network. Modern Pathol 8: 162A, 1995 (Abstract).

12) Swets J.A.: Measuring the accuracy of diagnostic systems. Scinces 240: 1285-1293, 1988.

13) Weinstein R.S., Bloom K.J., Rozek S.: telepathology and the networking of pathology diagnostic services. Arch Pathol Lab Med 111: 646-652, 1987.